UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

AD-A197 231

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
REPORT NUMBER TR42	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
TITLE *(and Subtitle)* Wavelength or Sensor Selection by Minimization of Prediction Error		5. TYPE OF REPORT & PERIOD COVERED Technical Report - Interim 6. PERFORMING ORG. REPORT NUMBER
AUTHOR(s) Kenneth R. Beebe and Bruce R. Kowalski		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0536
PERFORMING ORGANIZATION NAME AND ADDRESS Laboratory for Chemometrics Department of Chemistry BG-10 University of Washington		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 051-565
CONTROLLING OFFICE NAME AND ADDRESS Materials Sciences Division Office of Naval Research		12. REPORT DATE July 1, 1988 13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

Submitted and accepted for publication in ANALYTICAL CHEMISTRY.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Multivariate Calibration
Chemometrics
Sensor Selection
Wavelength Selection

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

A method of variable selection based on prediction error is presented. An algorithm based on this approach is developed and used to select variables using a simulated data set. The results of this procedure are compared to those expected using signal averaging with a smaller set of variables.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

OFFICE OF NAVAL RESEARCH

Contract N00014-75-C-0536

Task No. NR 051-565

TECHNICAL REPORT NO. 42

Wavelength or Sensor Selection by Minimization of Prediction Error

by

Kenneth R. Beebe and Bruce R. Kowalski

Prepared for Publication
in
Analytical Chemistry

Department of Chemistry, BG-10
University of Washington
Seattle, Washington 98195

July 1, 1988

Sir: An option that is often considered when using methods from multivariate analysis for instrument calibration is that of deleting one or more of the measurement variables from the model. For example, of the absorbances measured at 1024 wavelengths (variables) in an infra-red experiment, the analyst may want to use a small subset of the wavelengths for quantitative analysis to simplify the data processing. Other motivations may be the selection of the optimal subset of possible physical sensors to include in an array, a reduction in the number components and therefore the cost of an instrument, or to follow one of the rationalizations in the literature involving variable selection (1-3). In all situations, the analyst is concerned with determining the best subset of sensors or wavelengths to retain.

It can be shown that the only sure way of determining the optimal subset of variables is to test all possible subsets (4). Unfortunately, this becomes an unreasonable task for even a moderate number of variables. Many schemes have been developed to aid the analyst in this area without resorting to an all possible subset analysis. These methods have been shown to yield results that closely approximate those obtained by the all possible subsets approach (5). They include procedures based on: Principal Components Analysis (6), stepwise regression and minimization of squared errors (7), Mallows' Cp statistic (7), the condition number of the design matrix (8), the branch and bounds approach (2,9), and other algorithms developed to approximate the all possible subset calculation for a given number of desired variables (10).

In all of these procedures, the selection procedure first defines an optimization criterion or objective function and then employs a searching algorithm to determine a subset of variables that yields the 'best' or near-best value for the objective function. The major emphasis of this paper is to suggest an alternative optimization criterion to that employed in the more common variable selection procedures. Additionally, the results of a variable selection algorithm applied to simulated data will be used to discuss a common

misconception concerning the advantage of signal averaging over the use of additional unique variables.

The typical problem in analytical chemistry involves a two step procedure of modelling (calibration) and prediction (analysis of an unknown sample). The first step involves recording the instrument responses to a set of calibration samples. A model is then derived that describes the relationship between these responses and the concentrations of analytes in the samples. An accepted measure of the inability of the model to fit the data (lack of fit) is the residual sum of squared errors (RSS). This is the sum of the squared differences between the actual concentrations of the calibration samples and the model's estimates of these values where a weighted sum can be used where appropriate.

A common practice is to assume that the "best" model is that which minimizes RSS. Although this seems reasonable, and may be for some applications, the analyst should be aware of the implication of this approach. RSS is a measure of ability of the model to fit the calibration data and may or may not be a reasonable measure of the predictive ability of a model. It is important to understand the distinction between fitting and predictive ability. A very flexible calibration method that consistently yields models with low RSS values for any set of calibration samples will also perform well at fitting outliers (samples that are not representative of expected future samples). This phenomenon is termed overfitting and results in models that "flex" to describe outliers by compromising its ability to describe the true model parameters. In this case, the flexibility that increases model fit (reduces RSS) does so at the expense of predictive ability. An example of this type of calibration procedure is one that forces the model to pass through every calibration point. This procedure would yield RSS=0.0 for every calibration set used. In this extreme situation it is easy to see how predictive ability would be decreased if non-representative samples were used in the calibration step. It illustrates the danger of ignoring prediction when setting up an analysis where the ultimate goal is prediction. This

is true also in the context of variable selection.

Note that this does not mean that selection procedures based on optimization of prediction will work well in spite of poorly chosen calibration samples. The point is that RSS is not measuring the correct criterion. RSS should not be used as the selection criterion in situations where it is not a good measure of predictive ability.

In the sections that follow, an algorithm for variable selection is presented where minimization of prediction error (11) is used as the optimization criterion. The method assumes a knowledge of the noise structure of the response and sensitivity matrices and uses computer simulations to perform the selection. An alternative method is briefly presented in situations where the noise structure is not known. The algorithm is not meant to be the definitive answer to variable selection by minimization of prediction error, but rather an example of this alternative view in action.

Method

As the method is based on simulations, it is necessary to first present the mathematical formulation of a multivariate analysis to make clear the steps of the simulations. In all equations matrices are represented by bold uppercase letters (**C**); vectors by bold lowercase letters (**c**); and scalars by plain upper and lower case letters (I, i). All programs were written in FORTRAN 77 programming language and executed on a MicroVAX II computer. The discussions that follow are in the context of variable selection with the understanding that this is synonymous with sensor or wavelength selection.

Calibration. The typical analysis begins by measuring the instrument response at J variables to K analytes in I calibration samples having known analyte concentrations. A

model relating the responses to the concentrations is derived that satisfies the following relationship,

$$\mathbf{R} = \mathbf{C}\,\mathbf{S} + \mathbf{E} \qquad 1)$$

R is an I x J matrix of responses with J columns corresponding to the instrument response to I samples; **C** is the I x K concentration matrix of the I samples (rows) and K analytes; **S** is a K x J matrix of instrument sensitivities to the analytes, with the sensitivities at the J variables corresponding to columns; and **E** is an I x J matrix of errors that describes the inability of **S** to perfectly model the relationship between **R** and **C**.

Linear algebra can be used to determine the elements of the matrix **S** that minimize the sum of squares of the terms in **E** in the following manner (7),

$$\mathbf{S} = (\mathbf{C}^T\mathbf{C})^{-1}\,\mathbf{C}^T\,\mathbf{R} \qquad 2)$$

Where the matrix $(\mathbf{C}^T\mathbf{C})^{-1}$ represents the inverse of $\mathbf{C}^T\mathbf{C}$ and $(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$ is called the generalized inverse of **C**.

If the analyst follows some experimental design scheme, this inverse is simple to obtain. In situations where this inverse is difficult or impossible to obtain (where $\mathbf{C}^T\mathbf{C}$ is singular or near-singular), it is possible to calculate what is termed the *pseudo-inverse* of **C**, denoted as $\mathbf{C}^+$ (12). This is accomplished by using singular value decomposition (13) to decompose **C** into three matrices (**U**, **Σ**, and **V**) such that,

$$\mathbf{C} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T \qquad 3)$$

The pseudo-inverse of **C** can be obtained as,

$$\mathbf{C}^{+} = (\mathbf{V}^{T})^{-1}\,\Sigma^{-1}\,\mathbf{U}^{-1} \qquad 4)$$

The columns of **V** and **U** are mutually orthogonal, therefore, $\mathbf{V}^{T} = \mathbf{V}^{-1}$ and $\mathbf{U}^{T} = \mathbf{U}^{-1}$. This can be used to simplify equation 4 as follows,

$$\mathbf{C}^{+} = \mathbf{V}\,\Sigma^{-1}\,\mathbf{U}^{T} \qquad 5)$$

If one or more of the columns of the original **C** matrix are linear combinations of the remaining columns, some of the diagonal elements of Σ (the singular values) will approach zero and the inverse of these elements as found in equation 5 will approach infinity. To avoid this, one can retain only the Q largest diagonal elements of Σ while deleting the K - Q relatively small singular values. The K - Q columns of **V** and rows of $\mathbf{U}^{T}$ that correspond to these small singular values are also omitted. This results in the formation of new matrices $\mathbf{V}_Q$, Σ_Q^{-1}, and $\mathbf{U}_Q^{T}$ from which a close approximation to the pseudo-inverse of **C** can be calculated as follows,

$$\mathbf{C}^{+} \approx \mathbf{V}_Q\,\Sigma_Q^{-1}\,\mathbf{U}_Q^{T} \qquad 6)$$

Prediction. To perform prediction, it is necessary to calculate the inverse of the **S** matrix described in equation 2. This inverse can then be used to rearrange equation 1 to yield,

$$\mathbf{R}\,\mathbf{S}^{+} = \mathbf{C} \qquad 7)$$

Again the pseudo-inverse of **S**, $\mathbf{S}^{+}$, is used in place of $\mathbf{S}^{T}(\mathbf{S}\mathbf{S}^{T})^{-1}$ to make the

procedure useful in situations where the latter inverse is not obtainable. The analysis of an unknown sample (r_{un}) is then performed using the following equation,

$$r_{un} S^{+} = c_{un} \qquad 8)$$

Where S^{+} is the same pseudo-inverse found in equation 7 and c_{un} is the estimated concentration of the analytes in the unknown sample.

If the actual analyte concentrations of the "unknown" sample are known, such as when one calibration sample is withheld from the calibration step, a prediction step can be used to *validate* the calibration model. The sample r_{un} is termed a test sample and the error in the predicted concentration is used as an indication of the predictive ability of the model. This prediction error is termed a PRESS value (Predictive Residual Error Sums of Squares) and is calculated as follows,

$$PRESS = \sum_{k=1}^{K} [\, (c_{true})_k - (c_{un})_k \,]^2 \qquad 9)$$

Where $(c_{true})_k$ are elements of the vector of actual concentrations of analytes in the test sample, and $(c_{un})_k$ are the model's estimate of these values. The test sample is not used to build the calibration model and therefore can be used to test the predictive ability of the model. This type of validation procedure is a key step in the selection procedure described herein.

Variable Selection

To select variables via a prediction error minimization scheme, the following procedure can be followed.

1) The first step is the only experimental phase of the analysis with all remaining steps being computer simulations. The analyst first estimates the K x J matrix of sensitivities **S** as defined by equations 1 and 2 using a set of calibration samples. It is advisable to use as many samples as possible to obtain as accurate an estimate of **S** as is economically feasible.

2) Delete the first column from the **S** matrix, label the resulting K x (J - 1) matrix $\mathbf{S}_{-1}$.

3) Add Gaussian distributed noise to $\mathbf{S}_{-1}$ to simulate the uncertainty that is involved in the calibration step of an analysis. Call this matrix $\mathbf{S'}_{-1}$.

The calibration samples that were used in the simulations in this paper contained only one analyte at a time with unit concentrations. The matrix **R** therefore was equal to **S** and it was reasonable to add noise to **S** to simulate the uncertainty in the estimation of the sensitivity matrix. This uncertainty represents the fact that a repetition of the calibration process would not yield the exact same **S** matrix. An algorithm based on a random number generator is used to add Gaussian noise to each element in the $\mathbf{S}_{-1}$ matrix. The noise added in the simulations discussed in this paper is absolute Gaussian noise that is proportional to the largest value encountered in each column. If, for example, 1% noise is to be added to the vector $\mathbf{r} = [\ 1\ \ 3\ \ 4\ \ 2\ ]$, the procedure was to add Gaussian noise with mean zero and standard deviation .04 to all four elements of **r**. However, in all instances where noise is added to a matrix the analyst has the flexibility to add the amount and type that is appropriate for the particular instrument under investigation.

4) Specify the concentrations of the analytes in an appropriate set of unknown sample(s) to be used as a test set. Call this matrix $\mathbf{C}_{test}$. The concentrations of the different analytes are chosen by the analyst to reflect the typical samples that will be

analyzed by the instrument. The analyst can also leave one or more calibration samples out and use them as test samples. Similarly, each sample can be left out one at a time yielding I test samples on I slightly different calibration models. A more detailed description of this approach will be presented later.

5) Calculate the "true" responses for the (J - 1) variables of the instrument to the test sample(s) using equation 1 and the $\mathbf{S}_{-1}$ matrix derived in steps 1 and 2.

$$\mathbf{R}_{test} = \mathbf{C}_{test}\ \mathbf{S}_{-1}$$

6) Add noise to the matrix $\mathbf{R}_{test}$ to simulate the noise due to the instrument. Call this matrix $\mathbf{R'}_{test}$.

7) Calculate the pseudo-inverse of $\mathbf{S'}_{-1}$, $(\mathbf{S'}_{-1})^{+}$, and use it to estimate the concentrations of the analytes in the test samples using equation 4.

$$\mathbf{C}_{pred} = \mathbf{R'}_{test}\,(\mathbf{S'}_{-1})^{+}$$

8) Calculate the predictive residual sums of squares (PRESS) in one of the following manners:

If the goal of variable selection is to optimize the instrument for the analysis of one analyte in the mixture, the PRESS value is calculated for only that one analyte as follows,

$$PRESS_N = \sum_{i=1}^{I} (c_{pred\ i,N} - c_{test\ i,N})^2$$

Where the summation over i for the Nth column of C results in the optimization of the

prediction of the analyte occupying the Nth column.

To optimize the analysis for all of the analytes present, PRESS is calculated for all of the columns of **C** as follows,

$$\text{PRESS} = \sum_{i=1}^{I} \sum_{k=1}^{K} (c_{\text{pred } i,k} - c_{\text{test } i,k})^2$$

The importance of different analytes can be expressed in a weighting term as desired.

9) Return to step 3 and redo steps 3-8 using the same type of noise structure. The result of many iterations (>300) is a simulation of replicate analyses of the test samples ($\mathbf{C}_{\text{test}}$). Adding all of the PRESS values will yield a number that is a measure of the ability of the matrix $\mathbf{S}_{-1}$ to model the data.

10) After the PRESS value is calculated for $\mathbf{S}_{-1}$, the process is repeated from step 2 but with the second variable deleted. This results in a PRESS value for the model when the second variable is deleted. This is then repeated for all of the J columns of **S**.

11) The analyst then deletes the variable whose omission yields the smallest PRESS value, redefines **S** as the original **S** without this variable, and returns to step 2 to determine the second variable to delete.

This algorithm is based on a simple step down procedure and therefore will not necessarily yield the best subset of variables. As stated earlier, testing all possible combinations of variables guarantees the optimal set but is unfeasible in many situations because of the extremely large number of possible combinations that arise in fairly simple chemical applications. The simple step down method was selected to illustrate one possible approach among many.

The merits of this method are that it takes into account the error that is associated with both the calibration (step 3) and the measurement (step 6) phases of an analysis and

is based on the minimization of the prediction error. It is therefore both realistic and has an appealing basis.

Furthermore, this method can be used to calculate the PRESS value associated with a sensitivity matrix derived by deleting variables using another selection method by following steps 3 - 9 of the algorithm using the derived sensitivity matrix as $\mathbf{S}_{-1}$. This will yield a PRESS value for the selected variables and a given set of test samples. The PRESS values obtained using the same test samples and different variables to form $\mathbf{S}_{-1}$ can be used to compare between methods.

The requirements for the use of this algorithm are that the system obey the model described in equation 1, and an estimate of the noise structure of **S** and **R** can be obtained. This information can be obtained either from experimentation or by some theoretical basis. The algorithm then allows the analyst to simulate many experiments to optimize the variable selection based on predictive ability.

If the system contains non-linearities not described by equation 1 or the noise structure is unknown, similar methods based on jackknife (14) or cross-validation (15) can be used when data from the analysis of many samples is available. To jackknife the data, one begins by deleting the first sample data from the response and concentration matrices and building the calibration model with the remaining I - 1 samples. The model is applied to the instrument responses to this sample to predict the concentration of analytes. To determine the predictive ability of the model, a PRESS value for sample one is calculated summed over the desired analytes (see step 8 of algorithm). The data for this sample are then returned to **R** and **C** and the process repeated until each of the I samples has been removed from the data one at a time. The PRESS values obtained with each sample left out can be summed to yield a total PRESS value for the model. To use this procedure for variable selection, the analyst begins by deleting the Jth variable from **R**. The remaining I x (J-1) **R** matrix is jackknifed to yield a total PRESS value for the matrix

without the Jth variable. The data for this variable are returned to **R** and the process is repeated by deleting each of the J variables, one at a time, and calculating a total PRESS for the resulting I x (J-1) matrices. The variable that corresponds to the smallest total PRESS value is permanently dropped from **R** and the process is repeated for deleting the second variable.

The disadvantage of this latter method is that it requires data collected from many samples to obtain reproducible results. However, it is suitable in situations where many analyses have been performed and where the distribution of noise is not well known.

Results and Discussion

The first application used simulated data where the sensitivity matrix was formed from three Gaussian curves for the instrument response to three analytes (fig. 1). In this example, there were three analytes in the analysis and 29 measurement variables, none of which were perfectly selective for any of the analytes under investigation. This data set was chosen because it was simple to study and represents a generic collinear data set with many degrees of overlap.

The algorithm presented above was used to delete variables while minimizing the PRESS value for analyte 2 (see step 8) at a 1% absolute noise level, and the test sample $\mathbf{c} = [1\ \ 1\ \ 1]$ was used to determine the best set of variables. Figure 2 is the resulting plot of the PRESS when 1000 simulations were performed as one variable was deleted at a time. The number of iterations used was arbitrarily chosen by weighting the accuracy and precision of the results against computer time. One simulation with 1000 replicates took approximately 20 hours of CPU time on the MicroVAX II computer where little effort was made in optimizing the FORTRAN code. This may seem to be a large amount of time but one must also consider that this type of variable selection procedure is normally performed

only once for a given system.

Figure 2 indicates a general increase in PRESS as the variables are deleted one at a time. The plot is not perfectly smooth because of the noise that was added to the data and the finite number of iterations used for the simulations. As the number of iterations was increased, the plot became smoother. The main point is that increasing the number of variables improved prediction with the best predictions obtained by retaining all of the variables in the analysis.

It has been pointed out that using all of the variables decreases the precision of the analysis (1) because the inclusion of each variable adds noise to the analysis. However, the predictive ability of a model as measured by PRESS is a function of both the accuracy and precision, and figure 2 illustrates a situation where the accuracy increase realized by the use of all available variables compensates for the decreased precision. This is an important finding that was observed in all of the applications that have been studied using the present method.

Lorber and Kowalski (16,17) have presented a proof supporting this finding. They showed that an estimate of the variance of prediction for a model is given as,

$$\mathrm{Var}(c_{un}) \approx \sum_{j=1}^{J} b^2_{j,k}\ \mathrm{var}(r_{un,j}) + \kappa \sum_{i=1}^{I} h^2_{i,un}\ \mathrm{var}(c_{k,i}) \qquad 10)$$

where var(c_{un}) is the variance of the concentration estimates for an unknown sample; the $b_{j,k}$ are the regression coefficients for the model; $h_{i,un}$ are calculated as $\mathbf{h}^T_{un} = \mathbf{r}^T_{un}\,\mathbf{R}^+$; var($r_{un,j}$) and var($c_{k,i}$) are the variances of the unknown sample responses and the calibration concentrations, respectively; and κ is a scalar (16). From this equation, it was demonstrated that the addition of variables containing useful information to the model will always result in a decrease in prediction variance. Figures 2-4 substantiate these findings.

Although each of the 26 variables in these examples added noise to the calibration model, the best prediction (minimum PRESS) was realized using the full compliment of variables.

In figure 2, the final set of three variables consisted of numbers 3, 15, and 25. Comparing this result with figure 1 indicates it is a reasonable choice of variables with each analyte being represented by one variable. The "logical" choice of the most sensitive variable for each analyte was not realized due to the type of noise added and the fact that no three variables were definitively the "best" set. It was possible to derive a different final three variables that yielded a similar PRESS value. Again this is also a function of the number of iterations used for the simulations. As this number approaches infinity, theoretically one would expect one best set to emerge unless two or more sets are exactly equivalent.

Figure 3 is the resulting PRESS plot when the same analysis was repeated. The final variables chosen were 1, 13, and 28 and the predictive ability of the two different "best" sets were almost identical. It is important to realize that two completely unique sets of variables will often perform in a very similar manner. This is usually the case for spectra with broad features (e.g. NIR). A technique used to determine a subset will often yield one "best" set, but comparison of the predictive ability of this subset with that chosen using another method may yield very similar results that are statistically identical.

Once the smallest "best" subset of variables is obtained, the analyst may want to add additional information to improve the precision and accuracy of the prediction. One common approach is to measure the best variables more than once in order to decrease the measurement standard deviation. This procedure is termed signal averaging. Another option to consider is that of adding more unique variables to the analysis. An examination of figures 2 and 3 reveals that given the present system, the addition of unique variables is more beneficial than signal averaging. In figure 2, as the number of variables is increased from three to four (26 variables deleted to 25 deleted), a resulting decrease in the PRESS

value from 5.5×10^{-4} to 3.0×10^{-4}, is observed. According to theory, to achieve the same increase in precision using signal averaging would require 3-4 replicates at each of the three variables. It is clear from these examples that for this system, using more variables is a more efficient means of improving prediction.

To further study the appropriateness of signal averaging in this particular situation, the selection algorithm was modified to perform a step up selection (adding variables while minimizing PRESS). Optimizing again for the second analyte, the program began with the best set of three variables derived in the first experiment (variables 3, 15, and 25) and added the variables that resulted in minimum PRESS values. To allow for signal averaging, the algorithm was allowed to select a variable more than once. For example, if the addition of variable 3 to the original set (3,15,25) resulted in the smallest PRESS, it would be chosen even though it is already present and the new subset would consist of (3,15,25,3). This is equivalent to signal averaging on variable 3 using two independent measurements.

If signal averaging is to be preferred over the addition of unique variables, one would expect the step up procedure to select the variables that are already in the best minimum subset. The PRESS results for the addition of 26 variables to the original three are given in figure 4 and the order of variable addition was as follows: [(3, 15, 25), 16, 13, 12, 1, 9, 1, 12, 14, 12, 12, 20, 14, 12, 11, 12, 28, 7, 2, 29, 18, 24, 6, 4, 27, 3, and 18]. The method chose 20 unique variables and did not signal average (select a variable already chosen) until the ninth variable where variable one was reused. Again, it is important to stress that the decrease in PRESS realized by the addition of unique variables simply could not have been achieved using the precision improvement obtained by signal averaging alone. Signal averaging becomes a reasonable choice only when the amount of unique information contained by the remaining unique variables is small.

These results and those obtained by Lorber and Kowalski demonstrate that signal

averaging is not generally a substitute for multichannel detection even in instances where the number of total measurements is fixed. Where optimal prediction is sought, the value of adding a sensor to a given set is a function of the unique information it brings to the array as well as its signal to noise ratio.

Literature Cited:

1) Frans, S.D.; Harris, J.M. *Anal. Chem.* **1985**, 57, 2680-2684.

2) Sasaki, K.; Kawata, S.; Minami, S. *Appl. Spectrosc.* **1986**, 40, 185-189.

3) Honigs, D.E.; Hieftje, G.M.; Hirschfeld, T. *Appl. Spectrosc.* **1984**, 38, 844-847.

4) Cover, T.M. *IEEE Trans. Syst. Man. Cybern.* **1974**, 4, 116-117.

5) Berk, K.N. *Technometrics* **1978**, 20, 1-6.

6) Carey, W.P.; Beebe, K.R.; Kowalski, B.R.; Illman, D.L.; Hirschfeld, T.; *Anal. Chem.* **1986**, 58, 149-153.

7) Draper, N.; Smith, H. *Applied Regression Analysis* 2nd ed., **1981**, Wiley: New York.

8) Ebel, S.; Glaser, E.; Abdulla, S.; Steffens, U.; Walter, V. *Fresnius' Z. Anal. Chem.* **1982**, 313, 24-27.

9) Furnival, G.M.; Wilson, R.B. *Technometrics* **1974**, 16, 399-512.

10) Lamotte, L.R.; Hocking, R.R. *Technometrics* **1970**, 12, 83-93.

11) Allen, D.M.; Technical Report no. 23, **1971**, Department of Statistics, University of Kentucky.

12) Strang, G.*Linear Algebra and its Applications* 2nd ed., **1980**, Academic Press: New York.

13) Golub, G.H.; Van Loan, C.F. *Matrix Computations* **1983**, John Hopkins University Press: Baltimore.

14) Mosteller, F.; Tukey, J.W. *Data Analysis and Regression* **1977**, Addison-Wesley: Reading.

15) Wold, S. *Technometrics* **1978**, 20, 397-405.

16) Lorber, A.; Kowalski, B.R. *J Chemometrics* **1988**, 2, 67-79.

17) Lorber, A.; Kowalski, B.R. *J Chemometrics* **1988**, 2, 93-109.

Figure Captions:

Figure 1. Plot showing the thirty variables' sensitivities to the three analytes.

Figure 2. Simulation number 1: Step down selection with 1000 iterations. PRESS value as the variables are deleted one at a time.

Figure 3. Simulation number 2: Step down selection with 1000 iterations. PRESS value as the variables are deleted one at a time.

Figure 4. PRESS value as one variable is added at a time. Step up selection with 1000 iterations and allowing for signal averaging.

Figure 1.

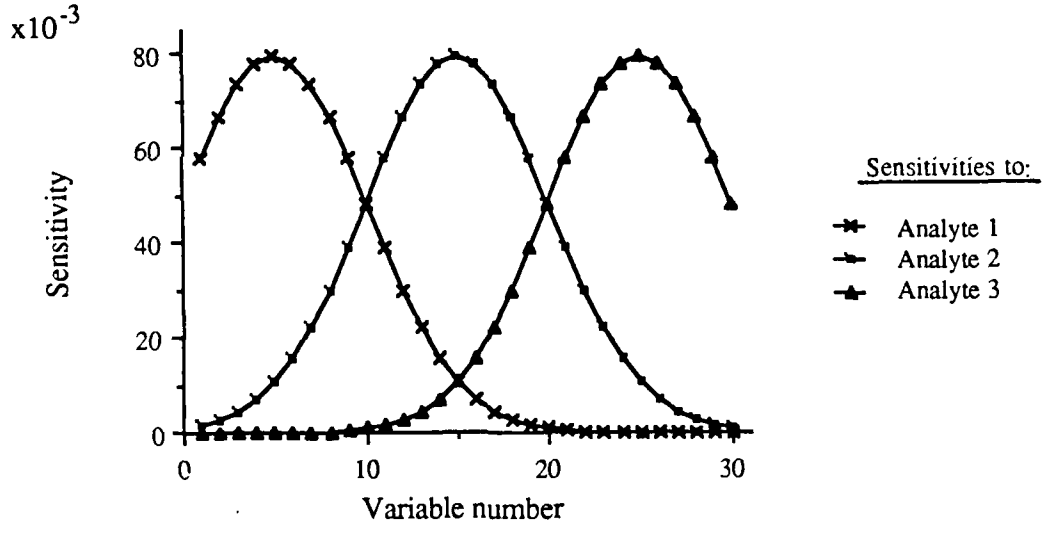

x10-3
80
60
40
20
0
Sensitivity
0
10
20
30
Variable number
Sensitivities to:
Analyte 1
Analyte 2
Analyte 3

Figure 2.

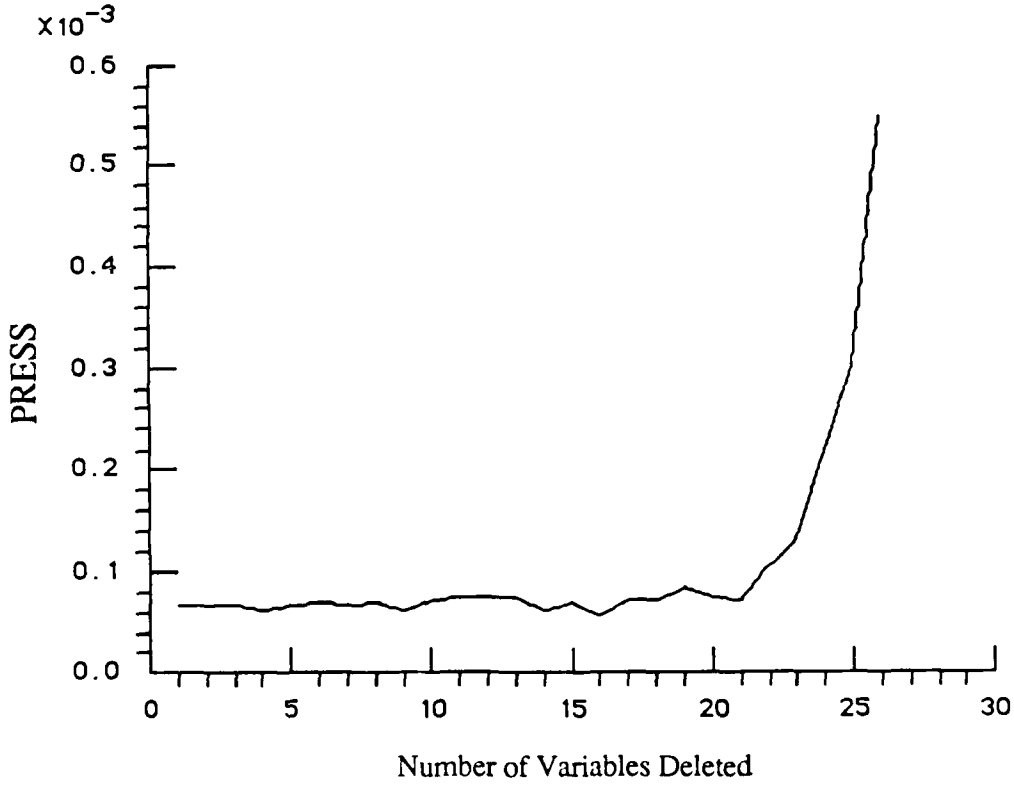

X10^{-3}
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PRESS
0
5
10
15
20
25
30
Number of Variables Deleted

Figure 3.

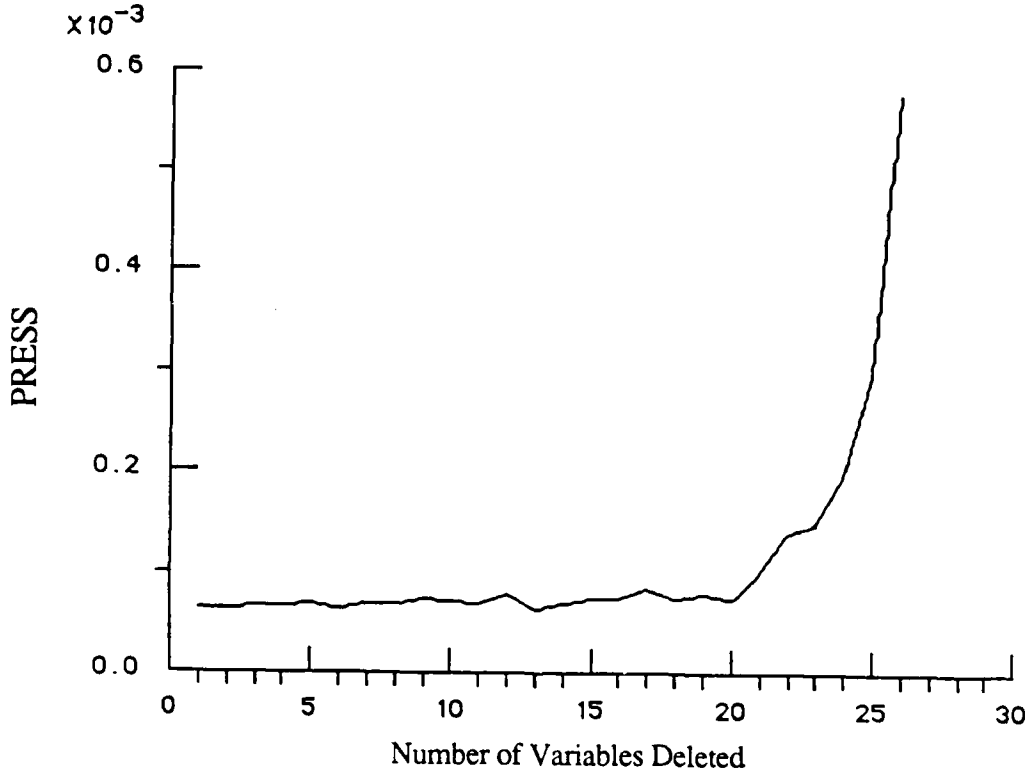
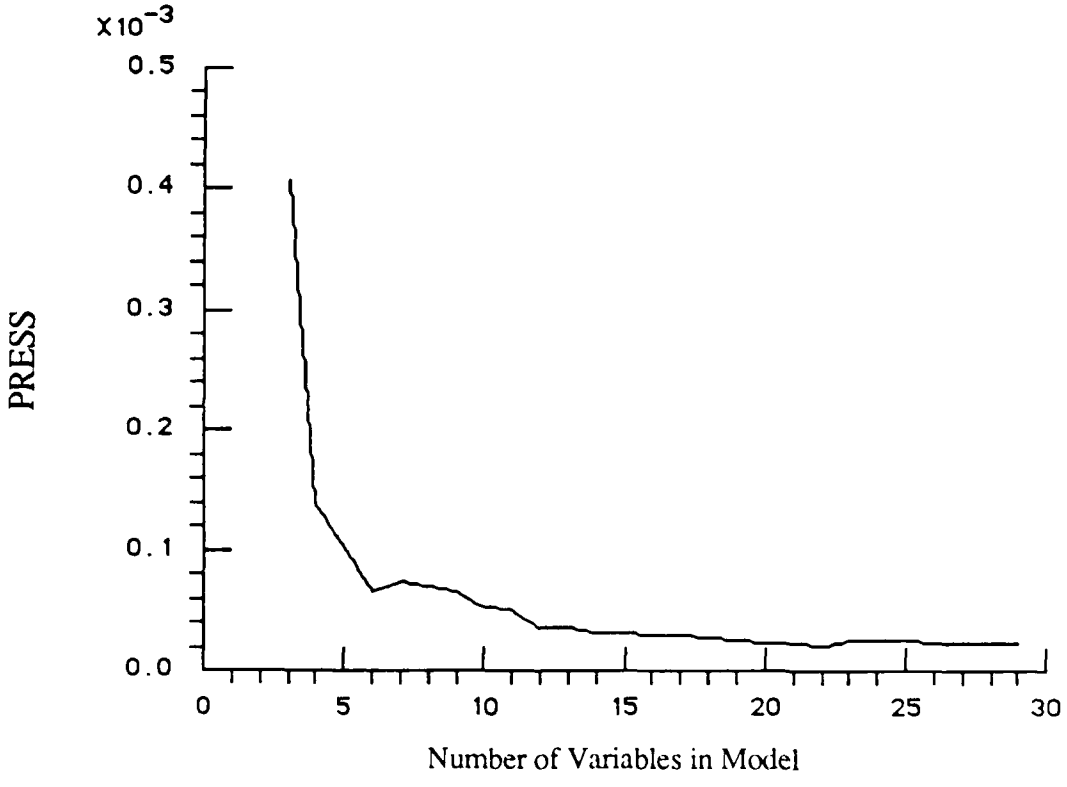

X 10⁻³
0.6
0.4
0.2
0.0
PRESS
0
5
10
15
20
25
30
Number of Variables Deleted

Figure 4.


X10^-3
0.5
0.4
0.3
0.2
0.1
0.0
PRESS
0
5
10
15
20
25
30
Number of Variables in Model

DL/413/83/01
GEN/413-2

TECHNICAL REPORT DISTRIBUTION LIST, GEN

	No. Copies
Office of Naval Research Attn: Code 413 800 N. Quincy Street Arlington, Virginia 22217	2
Dr. Bernard Douda Naval Weapons Support Center Code 5042 Crane, Indiana 47522	1
Commander, Naval Air Systems Command Attn: Code 310C (H. Rosenwasser) Washington, D.C. 20360	1
Naval Civil Engineering Laboratory Attn: Dr. R. W. Drisko Port Hueneme, California 93401	1
Defense Technical Information Center Building 5, Cameron Station Alexandria, Virginia 22314	12
DTNSRDC Attn: Dr. G. Bosmajian Applied Chemistry Division Annapolis, Maryland 21401	1
Dr. William Tolles Superintendent Chemistry Division, Code 6100 Naval Research Laboratory Washington, D.C. 20375	1
Dr. David Young Code 334 NORDA NSTL, Mississippi 39529	1
Naval Weapons Center Attn: Dr. A. B. Amster Chemistry Division China Lake, California 93555	1
Scientific Advisor Commandant of the Marine Corps Code RD-1 Washington, D.C. 20380	1
U.S. Army Research Office Attn: CRD-AA-IP P.O. Box 12211 Research Triangle Park, NC 27709	1
Mr. John Boyle Materials Branch Naval Ship Engineering Center Philadelphia, Pennsylvania 19112	1
Naval Ocean Systems Center Attn: Dr. S. Yamamoto Marine Sciences Division San Diego, California 91232	1

OL/413/83/01
051B/413-2

ABSTRACTS DISTRIBUTION LIST, 051B

Dr. R. A. Osteryoung
Department of Chemistry
State University of New York
Buffalo, New York 14214

Dr. J. Osteryoung
Department of Chemistry
State University of New York
Buffalo, New York 14214

Dr. H. Chernoff
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dr. A. Zirino
Naval Undersea Center
San Diego, California 92132

Dr. George H. Morrison
Department of Chemistry
Cornell University
Ithaca, New York 14853

Dr. Alan Bewick
Department of Chemistry
Southampton University
Southampton, Hampshire
ENGLAND SO9 5NH

Dr. M. B. Denton
Department of Chemistry
University of Arizona
Tucson, Arizona 85721

Dr. S. P. Perone
Lawrence Livermore National
Laboratory L-370
P.O. Box 808
Livermore, California 94550

Dr. G. M. Hieftje
Department of Chemistry
Indiana University
Bloomington, Indiana 47401

Dr. Christie G. Enke
Department of Chemistry
Michigan State University
East Lansing, Michigan 48824

Walter G. Cox, Code 3632
Naval Underwater Systems Center
Building 148
Newport, Rhode Island 02840

Professor Isiah M. Warner
Department of Chemistry
Emory University
Atlanta, Georgia 30322

Dr. Kent Eisentraut
Air Force Materials Laboratory
Wright-Patterson AFB, Ohio 45433

Dr. Adolph B. Amster
Chemistry Division
Naval Weapons Center
China Lake, California 93555

Dr. B. E. Douda
Chemical Sciences Branch
Code 50 C
Naval Weapons Support Center
Crane, Indiana 47322

Dr. John Eyler
Department of Chemistry
University of Florida
Gainesville, Florida 32611

DL/413/83/01
0518/413-2

ABSTRACTS DISTRIBUTION LIST, 051B

Professor J. Janata
Department of Bioengineering
University of Utah
Salt Lake City, Utah 84112

Dr. J. Dec[illegible]no
NAVSEA
Code 05R14
Washington, D.C. 20362

Dr. Charles Anderson
Analytical Chemistry Division
Athens Environmental Laboratory
College Station Road
Athens, Georgia 30613

Dr. Bob Flemming
B 108 Reactor
National Bureau of Standards
Washington, D.C. 20234

Dr. Frank Herr
Office of Naval Research
Code 422CB
800 N. Quincy Street
Arlington, Virginia 22217

Professor E. Keating
Department of Mechanical Engineering
U.S. Naval Academy
Annapolis, Maryland 21401

Dr. M. H. Miller
1133 Hampton Road
Route 4
U.S. Naval Academy
Annapolis, Maryland 21401

Dr. Clifford Spiegelman
National Bureau of Standards
Room A337 Bldg. 101
Washington, D.C. 20234

Dr. Denton Elliott
AFOSR/NC
Bolling AFB
Washington, D.C. 20362

Dr. B. E. Spielvogel
Inorganic and Analytical Branch
P.O. Box 12211
Research Triangle Park, NC 27709

Ms. Ann De Witt
Material Science Department
160 Fieldcrest Avenue
Raritan Center
Edison, New Jersey 08818

Dr. A. Harvey
Code 6110
Naval Research Laboratory
Washington, D.C. 20375

Mr. S. M. Hurley
Naval Facilities Engineering Command
Code 032P
200 Stovall Street
Alexandria, Virginia 22331

Ms. W. Parkhurst
Naval Surface Weapons Center
Code R33
Silver Spring, Maryland 20910

Dr. M. Robertson
Electrochemical Power Sources Division
Code 305
Naval Weapons Support Center
Crane, Indiana 47522

Dr. Andrew T. Zander PI204
Perkin-Elmer Corporation
901 Ethan Allen Highway/MS905
Ridgefield, Connecticut 06877

DL/413/83/01
051B/413-2

ABSTRACTS DISTRIBUTION LIST, 051B

Dr. Marvin Wilkerson
Naval Weapons Support Center
Code 30511
Crane, Indiana 47522

Dr. J. Wyatt
Naval Research Laboratory
Code 6110
Washington, D.C. 20375

Dr. J. MacDonald
Code 6110
Naval Research Laboratory
Washington, D.C. 20375

Dr. H. Wohltjen
Naval Research Laboratory
Code 6170
Washington, D.C. 20375

Dr. John Hoffsommer
Naval Surface Weapons Center
Building 30 Room 208
Silver Spring, Maryland 20910

Dr. Robert W. Shaw
U.S. Army Research Office
Box 12211
Research Triangle Park, NC 27709